Sensitivity of Middle Atmospheric Ozone to Solar Proton Events: A comparison Between a Climate Model and Satellites

Abstract

Energetic particle precipitation (EPP) impact on the middle atmospheric ozone chemistry potentially plays an important role in the connection between space weather and Earth's climate system. A variant of the Whole Atmosphere Community Climate Model (WACCM-D) implements a detailed set of ionospheric D-region chemistry instead of a simple parameterization used in the earlier WACCM versions. This allows WACCM-D to capture the impact of EPP in more detail. Here, we validate the ion chemistry of the WACCM-D by analyzing the middle atmospheric ozone response to the EPP forcing during well-known solar proton events (SPEs). We use a multi-satellite approach to derive the middle atmospheric sensitivity for the SPE forcing as a statistical relation between the solar proton flux and the consequent ozone change. An identical sensitivity analysis is carried out for the WACCM-D model results, enabling one-to-one comparison with the results derived from the satellite observations. Our results show a good agreement in the sensitivity between satellites and the WACCM-D for nighttime conditions. For daytime conditions, we find a good agreement for the satellite data sets that include the largest

SPEs (maximum proton flux ___ pfu). However, for those satellite data-sets with only minor and moderate SPEs, WACCM-D tends to underestimate the sensitivity in daytime conditions. In summary, the WACCM-D with its full ion chemistry and transportation demonstrates a realistic representation of the SPE sensitivity of ozone, and thus provides a conservative platform for long-term EPP impact studies.

Key Points

- Atmospheric ozone sensitivity to solar proton events is derived from multiple satellites and Whole Atmosphere Community Climate Model (WACCM-D) model

- A good agreement exists between nighttime data sets. For daytime, there is agreement only when largest, major solar proton events (SPEs) are included

- The results indicate WACCM-D likely provides a realistic and/or conservative estimate of ozone sensitivity to SPEs

Plain Language Summary

In this paper, we test the atmospheric chemistry included in the most recent version of a state of the art climate model. For this test, we use measurements from four satellite instruments to derive the atmospheric ozone sensitivity to solar proton forcing as a statistical relation. We derive the same relation from the climate model simulations and perform a one-to-one comparison with the satellite results. Our comparisons show a good agreement between the simulated and observed sensitivity for nighttime conditions. For daytime conditions, we find a good agreement only when the strongest proton events are included. If we consider only weak and moderate proton events, the climate model tends to underestimate the sensitivity compared to the satellites. We conclude that, overall, the chemistry included in the model provides a realistic representation of the atmospheric impact from protons. This is important for long-term simulations of the space weather influences on climate.

1 Introduction

Energetic Particle Precipitation (EPP) consists of highly energetic protons, electrons and neutrons that precipitate into the Earth's atmosphere from extra-terrestrial sources. Solar Proton Events (SPEs) is a type of EPP where a large quantity of protons are ejected from the Sun at high energies (10–300-MeV). The SPEs are sporadic but tend to occur more likely during solar maximum. In the near-Earth space, the energetic protons are

guided by the Earth's magnetic field to the polar regions, typical energies providing access to geomagnetic latitudes (Verronen et al., **2007**). Depending on the energy of the protons, they can penetrate into the upper and middle atmosphere, and have a significant direct impact at altitudes above 30 km (Jia et al., **2020**). The energetic protons cause ionization and dissociation in the middle atmosphere that ultimately lead to ozone reduction (Sinnhuber et al., **2012**).

The SPE impact on middle atmospheric odd oxygen, (+ O) has been studied since the November 1969 SPE (Weeks et al., **1972**), and is generally well known (Funke et al., **2011**; Jackman et al., **2001**; Seppälä

et al., **2004**). The SPE-driven depletion is mainly caused by the increased production of odd

hydrogen, (H + OH + O) and odd nitrogen, (N + NO +). The production occurs through ion chemistry with water cluster and molecular oxygen reactions (Verronen & Lehmann, **2013**). The

SPE impact increases at altitudes from 40 to 90 km in particular (Häkkilä et al., **2020**), with the peak of

production typically being near 70 km altitude. has a short chemical lifetime of the order of 0.1–1 days

below 80 km, but it can cause a large reduction in (up to) near the peak production altitude during

SPEs (Verronen et al., **2006**). The production of occurs through dissociation of molecular nitrogen and ion

chemistry (Sinnhuber et al., **2012**). The increased production of by SPEs leads to observable enhancements from 40 to 90 km (Seppälä et al., **2004**; Päivärinta et al., **2013**). Being mainly lost through photo-

dissociation by UV radiation, in the polar winter conditions has a longer chemical lifetime and can be transported downwards. However, the catalytic ozone depletion requires solar light to produce atomic oxygen

and is therefore not efficient in full polar night. The descent can take place during the entire polar night, but the stratospheric impact is largest in late winter and spring (López-Puertas et al., **2005**; Päivärinta et al., **2016**; Seppälä et al., **2007**).

Whole Atmospheric Community Climate Model (WACCM) is a general circulation model developed by the National Center For Atmospheric Research (NCAR). It features atmospheric dynamics fully coupled with chemistry to simulate responses to radiation and geomagnetic forcing (Gettelman et al., **2019**; Marsh et al., **2007**, **2013**). A recently developed variant, WACCM-D, features a set of photo-chemistry equations for the ionospheric D-region which replaces a simple parameterization used before, thus WACCM-D has an improved representation of the atmospheric impact by EPP. A description and overview of the WACCM-D model is given by Verronen et al. (**2016**). To demonstrate WACCM-D modeling capabilities, Andersson

et al. (**2016**) studied the January 2005 SPE and the response in atmospheric   , active chlorine   ,

and   , and found a good agreement with satellite-based observations. Recently, a statistical SPE study using

Superposed Epoch Analysis of the WACCM-D results found a larger   reduction at 0.5 hPa due to

improved   response, and a lower   reduction near 0.01 hPa due to lower   production (Kalakoski

et al., **2020**). In these studies, the inclusion of ion chemistry in the model also led to increased   production. Using the same WACCM-D results as Kalakoski et al. (**2020**), a comparison of SPE impacts against a single satellite data set (MLS) was performed by Jia et al. (**2020**) and a good consistency was found in the spatial and temporal extent of the ozone response.

In this study, we use a multi-satellite approach to validate the capability of WACCM-D to simulate the ion chemistry for SPEs. We derive atmospheric sensitivity to the SPE forcing as a statistical relation between solar proton flux and the consequent ozone change at different altitudes using the satellites. An identical analysis for WACCM-D is carried out to conduct a one-to-one comparison with the satellites. It is worth noting, that this study evaluates the model's response to EPP in general, not only to the SPEs, as the middle-atmospheric ozone response to a certain amount of ionization is identical whether it is caused by protons or electrons. If the model reproduces the observed ozone levels successfully in the case of SPEs, it builds confidence on the models capability to capture the impact of electron precipitation correctly as well, given that its flux is known.

2 Data and Model Description

We use satellite-based ozone observations from the Sounding of the Atmosphere using Broadband Emission Radiometry (SABER), Rong et al. (**2009**), Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), Fischer et al. (**2008**), Microwave Limb Scanner (MLS), Waters et al. (**2006**) and Optical Spectrograph and Infra-Red Imager System (OSIRIS), Murtagh et al. (**2002**). They provide near global

coverage (latitude   south to latitude   north) with daytime and nighttime data, except OSIRIS only provides daytime. In addition, we use proton flux measurements from the Geostationary Operational Environmental Satellites (GOES), Rodriguez et al. (**2010**) and ozone data from WACCM-D. These satellites for ozone data were chosen as they have previously been used to investigate SPE impact on middle atmospheric ozone, see for example, for MIPAS (Funke et al., **2011**; López-Puertas et al., **2005**), OSIRIS (Degenstein et al., **2005**), MLS (Andersson et al., **2016**; Jia et al., **2020**; Verronen et al., **2006**) and SABER (Gao et al., **2017**; Päivärinta et al., **2013**).

The GOES satellites carry an energetic particle sensor and have provided continuous solar proton flux measurements since 1974 (Rodriguez et al., **2010**). The $E \geq$ 10 MeV proton flux by GOES is a useful parameter as it provides a direct measure of SPE impact on mesosphere and stratosphere where the SPE-driven ozone depletion takes place (Jackman & McPeters, **2004**; Seppälä et al., **2004**, **2006**; Verronen et al., **2005**, **2006**). Additionally, GOES data are used to identify SPEs as time periods where the 5-minute integrated proton flux with $E \geq$ 10 MeV is greater than 10 particle flux units (pfu,   ). A list of all detected SPEs, including times of occurrence and maximum pfu, are listed by the National Oceanic and Atmospheric Administration (NOAA) at **https://umbra.nascom.nasa.gov/SEP/**. The latest recommended solar forcing data set for the Coupled Model Intercomparison Project (CMIP6) includes proton ionization based on the GOES observations (Matthes et al., **2017**). We used data from GOES 3, 9, and 11 satellites, and the data are available at **https://www.ngdc.noaa.gov/stp/satellite/goes/dataaccess.html**.

SABER onboard TIMED satellite features a 10 channel broadband infrared limb radiometer, launched in December 2001 (Rong et al., **2009**). It measures middle atmospheric ozone using the 9.6 μm band, which has been validated for 15–70 km altitude (Rong et al., **2009**). An advantage of SABER is its high vertical resolution ($\sim$1 km without degradation), which benefits from the infrared property of the instrument that allows a low noise detection (Rong et al., **2009**). In our work, we used the 9.6 μm ozone data version 2.0 which is available at **http://saber.gats-inc.com/**.

MLS is a microwave thermal emission scanner onboard the AURA satellite, launched in July 2004 (Waters et al., **2006**). It measures atmospheric ozone using the 240 GHz emission lines and provides a vertical resolution of 3–6 km altitude (increasing for higher altitudes). We used version 4.2 ozone data which are available at **https://mls.jpl.nasa.gov/**.

OSIRIS onboard ODIN satellite is a combined spectrograph and infrared imaging system, launched February 2001 (Murtagh et al., **2002**). It measures atmospheric ozone from spectral dispersed and scattered sunlight with a vertical resolution of 1.5 km. We used level 2 ozone data provided by the European Space Agency Climate Change initiative (ESA-CCI) available at this repository **https://doi.org/10.23729/b4558152-206b-4776-b9cc-ea52e8282678**.

MIPAS onboard ENVISAT (Environmental Satellite) is a limb radiance scanner in the mid-infrared spectra, launched in March 2002 (Fischer et al., **2008**). It measures middle atmospheric ozone emission with spectrometer followed by Fourier transformation with vertical resolution of $\sim$3 km. We used level 2 ozone data provided by ESA-CCI available at this repository **https://doi.org/10.23729/b4558152-206b-4776-b9cc-ea52e8282678**.

From the WACCM model family of component sets, we make use of SD-WACCM-D version 4 with pre-configured specified dynamics from the Modern-Era Retrospective analysis for Research and Applications (MERRA), Rienecker et al. (**2011**). This means that the meteorological fields at altitudes below 50 km are forced at every dynamical time step, while the model is fully interactive at altitudes above. The vertical coverage of this WACCM version are from surface level up to 140 km altitude (hPa) through 88 pressure levels. The latitude and longitude resolution is 1.   and 2.   , respectively. The model time-step is 30 min. The simulations were driven by historical solar forcing, including both radiation and EPP (Marsh et al., **2007**). The EPP included in this simulation are SPE, Kp-driven aurora and Galactic Cosmic Rays. Because high proton

fluxes contaminate satellite-based electron flux observations during SPEs (Rodger et al., **2010b**), we do not

include ionization from Medium-Energy Electrons (MEE),   keV in our simulations. In any case, below 90 km altitude we expect that major SPEs will dominate over MEE in both magnitude as well as altitude and latitude extent of short-term atmospheric response (Häkkilä et al., **2020**). We further discuss the omission of MEE and justify our approach in Section **5**. We make use of the daily zonal average SPE input recommended for the CMIP6 simulations (Matthes et al., **2017**). These SPE forcing data are based on the GOES energy-flux measurements (Jackman et al., **2011**). GOES data provide well-defined and accurate forcing for simulations of atmospheric response (Funke et al., **2011**). The same WACCM setup was previously used by Kalakoski et al. (**2020**) and Jia et al. (**2020**) and thus more details can be found there. For our sensitivity analysis, we have selected ozone profiles from WACCM simulation results to have a one-to-one correspondence with satellite observations in terms of time, latitude, longitude, and altitude. Because we are analyzing a large number of data points statistically, we expect any differences caused by the finite resolution of the model to average out in the analysis. With this data selection, there is an equal amount of ozone profiles from satellites and SD-WACCM-D, that is, we have a simulated data set to match each of the satellite instruments.

3 Methods

In our sensitivity analysis, we use ozone profiles from the satellites and the WACCM-D at geographic latitudes

of   North and South. The SPE impact on ozone are typically seen in these polar regions as shown by satellite-based observations, for example, (Degenstein et al., **2005**; López-Puertas et al., **2005**; Seppälä et al., **2006**).

From the NOAA SPE list, we select SPEs between the years 2001–2012 that are available to SABER, OSIRIS, MIPAS and MLS, separately. For each SPE, we create a time series of ozone profiles from 24-h before the onset to 24-h after the time of maximum pfu. By concentrating the time series around the maximum pfu from SPEs allows us to focus on the maximum SPE impact on ozone. Additionally, using a short time period in general for the time series allows us to avoid overlapping SPEs and reduce variability caused by the atmospheric dynamics. Furthermore, we define the profiles recorded before SPE onset as reference or undisturbed ozone profiles, and the profiles recorded after SPE onset as disturbed ozone profiles. To ensure a reasonable observation of the SPE affected ozone level from the time series, we require them to have at least 10 undisturbed and 5 disturbed profiles. We have also separated the data sets into groups with daytime (solar zenith

angle lower than   ) and nighttime (solar zenith angle greater than   ) conditions to separate the data set with and without solar illumination.

The relative ozone change in percentage,   is calculated for each altitude separately from

$$\tag{1}$$

where   is an individual disturbed profile recorded after SPE onset. The avg[   ] is the average of all undisturbed or reference profiles recorded before SPE onset.

The ozone loss due to SPEs occurs through chemical reactions with   and   , and requires solar radiation. Hence, we expect a time delay between the atmospheric ionization by SPE and the corresponding ozone loss effect that depends also on local time, (Verronen et al., 2005). Therefore, for each of the relative ozone change

profiles we calculated the average proton flux with E   10 MeV over a 12-h time period (henceforward called proton forcing) prior to the disturbed ozone measurement time. Based on earlier studies see for example, Seppälä et al. (2004); Degenstein et al. (2005); Rohen et al. (2005); López-Puertas et al. (2005); Verronen

et al. (2005, 2006); Seppälä et al. (2006); Jackman and McPeters (2004), we expect to see SPE-driven ozone depletion that correlates with the corresponding proton forcing. Here we derive a simple linear relation between the forcing and the ozone change

$$(2)$$

where   is an offset term (expected to be   0), and   the ozone sensitivity corresponding to the proton forcing   .

The sensitivity,   describes the ozone response to SPEs at an altitude. For example, negative

sensitivity,   indicates ozone depletion, while positive,   indicates increased production by SPEs. A

third possibility is   , which indicates there is no ozone response to the SPEs at this altitude. For the error

analysis, the 95% confidence intervals of the sensitivity,   are estimated using the following equations

$$(3)$$

where   is the Student's t-distribution with   degrees of freedom, N is the number of data samples used,

RMSE is the Root Mean Squares Error of the fit and   is the average of proton forcing,   (Walpole et al., 2012). Deriving the sensitivity parameter and its uncertainty identically for both the satellite data-sets and the corresponding WACCM-D profiles, enables a direct experimental comparison to validate the WACCM-D response to SPE-driven ionization.

4 Results

To illustrate the sensitivity analysis with the measurements, Figure **1** shows the relative ozone change as a function of proton forcing from a collection of SPEs measured by SABER. The plot is shown in a semi-logarithmic *x*-axis scale to better view the distribution of the data.

Figure 1
Open in figure viewerPowerPoint
Relative ozone change (in %) during daytime conditions at 70 km altitude as function of proton forcing (12-h time averaged proton flux, pfu). (a) Sounding of the Atmosphere using Broadband Emission Radiometry (SABER) observations in Northern Hemisphere (NH); (b) Whole Atmosphere Community Climate Model (WACCM-D) simulations in NH; (c) SABER observations in Southern Hemisphere (SH); (d) WACCM-D simulations in SH. Each dot represents an individual ozone measurement and the related proton forcing. The color separates the data points according to the individual SPEs. The solid black line is a least squares linear model fitted to the individual dots and the dashed black line is the RMSE of the model. The

coefficients   and   are the offset and sensitivity, respectively. Their uncertainties are 95% confidence intervals. The *y*-axis of these panels have been set from −100% to 100%, hence there are 1 missing dot in panel (b) near 4 pfu at 325% and two missing dots in panel (c) near 2 pfu at 220% and 1100%.

Here we consider the daytime data from both hemispheres and compare the SABER observations with the co-located WACCM-D simulations. Note that NH (Northern Hemisphere,   ) and SH (Southern Hemisphere,   ) have a different sets of SPEs available from the SABER observations.

As seen from Figure **1**, the SPEs in this set are distributed over a proton forcing range between   to   proton forcing pfu. We observe the ozone change to vary around zero along with low-flux SPEs, indicating no significant response. Above a certain flux limit (typically between   pfu) the decrease of ozone becomes evident and exceeds the background variability. Compared to the ozone decrease observed in SH by SABER, the simulated response from WACCM-D tends to be less sensitive to SPEs.

Figure **1** also shows the linear model (Equation. **2**) fitted to the data and the sensitivity, that is, slope of the fitted lines. In this case, the comparison reveals that SABER is   1.9 and   2.5 times more sensitive to SPEs than WACCM-D in the NH and SH, respectively. A large number of the individual data points are at low pfu below an apparent threshold pfu for any significant ozone response. To test the robustness of our fit, we removed data points with proton forcing less than 10 pfu, repeated the fit and compared it to our original result (figure not shown). Results reveal the RMSE of the fits and the uncertainty of the fit coefficients to be over-lapping, that is, not significant different. This is an indication of the robustness of our approach.

To extend the analysis to all altitudes, and to include both daytime and nighttime conditions, Figures **2** and **3** represent the sensitivity coefficients with their uncertainties for SABER, MLS, MIPAS, and OSIRIS, plotted against the corresponding WACCM-D results. Note that each panel represents a different set of condition to compare WACCM-D against satellite observations, that is, different subsets of SPEs, photo-chemical conditions (daytime or nighttime), locations, and seasons. However, the comparisons between the WACCM-D and the satellite data inside any single panel represents an identical subset of conditions, and therefore should reveal the true differences in the sensitivity. For example, Figures **2b** and **2d** shows a different shape of the sensitivity profiles from the two instruments, while WACCM-D shows the same difference and agrees with both instruments. The multi-satellite approach is used here to cover the maximal range of conditions in testing the model against the observations. Table **1** lists the number of SPEs collected in each panel, and furthermore, the largest SPE event (in terms of pfu) of the each set.

Figure 2
Open in figure viewerPowerPoint

Relative ozone change sensitivity coefficient,   with 95% confidence intervals as uncertainty from linear model as function of altitude in the northern polar region. Blue line represents satellite while red line represents Whole Atmosphere Community Climate Model (WACCM-D). (a) Sounding of the Atmosphere using Broadband Emission Radiometry (SABER) during daytime; (b) SABER during nighttime; (c) Microwave Limb Scanner (MLS) during daytime; (d) MLS during nighttime; (e) Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) during daytime; (f) MIPAS during nighttime; (g) Optical Spectrograph and Infra-Red Imager System during daytime. Note that we have set limits on the *x*-axis, so there are measurements in (c) and (g) outside the limits.

Figure 3
Open in figure viewerPowerPoint

The same as previous figure, but for south polar region. Note that we have set limits on the *x*-axis, so there are measurements in (g) outside the limits.

Table 1. Largest SPE as Maximum Proton Flux and Number of Available SPE Given in Pfu Ranges for Figures 2 and 3

Satellite, panel	Largest pfu	10 pfu 100	100 pfu 1000	1000 pfu 10,000	pfu 10,000
SABER, a.)	6,530	13	2	3	0
MLS, c.)	6,530	21	10	7	0
MIPAS, e.)	29,500	16	8	4	1
OSIRIS, g.)	2,520	16	5	3	0
North polar region, during daytime condition (Figure 2, left panel)					
Satellite, panel	Largest pfu	10 pfu 100	100 pfu 1000	1000 pfu 10,000	pfu 10,000
SABER, b.)	29,500	9	7	5	1
MLS, d.)	6,530	8	5	4	0
MIPAS, f.)	29,500	6	5	4	1
North polar region during nighttime condition (Figure 2, right panel)					
Satellite, panel	Largest pfu	10 pfu 100	100 pfu 1000	1000 pfu 10,000	pfu 10,000
SABER, a.)	1980	8	4	2	0
MLS, c.)	6,530	19	9	7	0
MIPAS, e.)	29,500	11	7	4	1
OSIRIS, g.)	31,700	8	5	2	2
South polar region during daytime condition (Figure 3, left panel).					
Satellite, panel	Largest pfu	10 pfu 100	100 pfu 1000	1000 pfu 10,000	pfu 10,000

Satellite, panel	Largest pfu	10 pfu 100	100 pfu 1000	1000 pfu 10,000	pfu 10,000
SABER, b.)	2,520	15	7	2	0
MLS, d.)	6,530	19	7	4	0
MIPAS, f.)	6,530	12	5	1	0

South polar region during nighttime condition (Figure 3, right panel)

In general, the results shown in Figures **2** and **3** reveals sensitivity and its uncertainty being negative above circa 40 km, and decreasing with altitude until a minimum at 65–70 km. This indicates a statistically significant SPE impact on ozone in within these altitudes. SABER, which has altitude coverage above 70 km, shows a relaxation above the peak impact towards no effect at 80 km altitude. A positive NH impact is observed in daytime around 40–45 km by SABER, while MLS observes this impact both during daytime and nighttime. In the SH only MLS observes a similar increase but during nighttime.

Overall comparison of the ozone sensitivity between satellites and WACCM-D shows a good agreement, especially in nighttime. In daytime, however, most of the comparisons suggest that the WACCM-D underestimates the SPE impact, with possible exceptions of MIPAS (in both NH and SH), and OSIRIS in SH.

5 Discussion

In general, our results show a clear depletion of ozone by the SPEs between 40–80 km with the largest impact around 70 km in altitude. This corresponds well to the ozone reduction seen in earlier case studies of SPEs. For example, studies on the 2003 October SPE impact on ozone by Seppälä et al. (**2004**); Verronen et al. (**2005**); López-Puertas et al. (**2005**); Degenstein et al. (**2005**), the January 2005 SPE impact on ozone by Verronen et al. (**2006**); Seppälä et al. (**2006**); Andersson et al. (**2016**), and recent statistical studies by Kalakoski et al. (**2020**); Jia et al. (**2020**) have shown similar results.

In some cases, also an enhancement of ozone may occur during an SPE, see for example, Figure **2c** at 45 km altitude. Similar net increase of ozone near this altitude was reported by Verronen et al. (**2006**) when they investigated the January 2005 SPE, featuring a 5040 maximum proton flux. Two possible contributing factors for this positive anomaly are pointed out by Verronen et al. (**2006**). First, it could be the so-called ozone self-healing effect, in which ozone decrease at an altitude (e.g., by SPE) allows more UV-driven ozone production at altitudes below under certain conditions (Jackman & McPeters, **1985**). Second, it could be related to dynamical transport in the horizontal direction. Note that WACCM-D has both chemical and dynamical representation, which seems to enable WACCM-D to produce the positive sensitivity in agreement with satellite-observations.

As a summary, the comparison of the ozone sensitivity to SPEs between satellites and WACCM-D shows, in general, a good agreement, especially in the nighttime conditions for both NH and SH. For daytime condition, we only see agreement in three of the eight comparisons: two in SH (MIPAS, OSIRIS) and one in NH (MIPAS). From Table **1**, we see that these three comparisons include also large SPEs (maximum proton

flux   pfu) in their data sets. The other 5 comparisons showing disagreement are limited to moderate and

small SPEs (maximum proton flux   pfu). This suggests the agreement in sensitivity comparison between

satellites and WACCM-D are related to the magnitude distribution of the SPEs included in the set. Therefore, we performed a repeated analysis for all the WACCM-D versus data comparisons after restricting them to moderate and small SPEs (figures not shown).

For nighttime conditions, we have SABER in NH and MIPAS in SH. Restricting these data sets by excluding the largest SPEs, we continue to see a general good agreement. For daytime conditions, we have MIPAS data in NH and SH, and OSIRIS data in SH. Restricting these data sets by excluding the largest SPEs, OSIRIS and MIPAS become more sensitive than WACCM-D above 45 and 60 km altitude in SH and NH, respectively. This suggests that WACCM-D and its ion chemistry provides a conservative estimate on the ozone impact when only moderate and small SPEs are considered. We have not seen any cases where WACCM-D would be clearly more sensitive than the satellites.

The restriction of the data to include only small and moderate SPEs also shows other interesting changes in the results. For MIPAS and SABER during nighttime in NH we now observe a positive sensitivity at 40 and 44 km altitude, respectively. This is similar to what we see in Figures **2c** and **2d** at 45 km altitude. It is likely the positive response are derived from the mid-range SPEs. Our method of estimating the linear fit through all ozone change samples as function of proton forcing see the positive contribution from the mid-range SPEs as noise when large SPEs are included. When the large SPEs are excluded, the mid-range SPEs positive contribution becomes more prominent and is therefore seen.

We also observed the sensitivity for WACCM-D and satellite at the altitude of largest impact around 70 km in

altitude to be increased by a factor of   2.5 when restricting the data to small and moderate SPEs. This can be explained by considering the ozone change and the related proton forcing from the moderate and large SPEs. A large SPE, for example, the October 2003 SPE featuring a 29,500 max proton flux, has been reported to

reduce ozone by   (Verronen et al., **2005**). However, a moderate SPE, for example, January 2005 SPE

featuring a 5040 max proton flux, has been reported to reduce ozone by   (Seppälä et al., **2006**). From

moderate to large SPEs, the proton flux increases by a factor   6, which yields a ozone change factor

of   1.4. This is reflected in our analysis method. As the difference in ozone change between moderate and large SPEs are not large compared to the difference in proton forcing, restricting our method to moderate SPEs will therefore significantly increase the magnitude of the negative sensitivity peak. However, this does not affect the one-to-one agreement between satellite and WACCM-D as the method is applied identically to comparable sets of data from the simulations and the observation. Further, we verified that the agreement between nighttime data sets remains even when restricting towards moderate SPEs.

The simulated data from WACCM-D uses daily average zonal mean SPE ionization rates identical to the recommended CMIP6 data set. These data are interpolated for each calculation time in between. This means the SPE forcing driving the simulations is smoothed compared to variability seen in shorter time scales, which could lead to observation-model differences in ozone response. In our method we combine a number of SPEs and create time series for each SPE up to 24-h after time of maximum pfu. In that time frame we may have an underestimation or overestimation of SPE forcing, depending on the case, but to some extend, these would then average out in the analysis. If the peak of an SPE occurred in the daytime when SPE forcing leads to more-or-less instantaneous catalytic chemistry and ozone depletion, then our simulations could be more prone to underestimate the ozone response and sensitivity. For nighttime, however, this is less likely due to nighttime response reflects a combination of catalytic loss and recovery from preceding daytime.

Proton precipitation has an energy-dependent geomagnetic cutoff latitude around (Heino & Partamies, **2020**; Heino et al., **2019**; Rodger et al., **2006**; Verronen et al., **2007**). In our method, we use ozone profiles sampled within geographic north and south latitude. However, the location of geomagnetic latitude varies between – geographic latitude along the geographic longitude. Additionally, in WACCM-D simulations proton precipitation is applied at and above geomagnetic latitude. On the other hand, the ozone response is dependent on and affected by atmospheric chemical and dynamical conditions, such that the extent of the impact as measured from satellites does not exactly follow the boundaries of SPE forcing. To test if this has an impact on our analysis, we combined new data sets from the satellites and the co-located WACCM-D by removing ozone profiles outside geomagnetic latitude (figure not shown). In general, we observe the same ozone response to SPEs, and the same agreement between satellite and WACCM-D as we do in Figures **2** and **3**. A comparison of the sensitivities from this modified data set to our original data set reveals the majority of the differences to be not statistical different (i.e., over-lapping uncertainties between data sets). The difference seen in the altitudes with a statistical significant difference are very small. This contributes to the fact we continue to see the same ozone response, the same agreement between satellite and WACCM-D, and therefore the overall conclusion remains the same.

As discussed in Section **2**, MEE ionization is not included in our simulations, which could in principle explain some of the model-observation differences. However, we argue that including MEE in our simulations would have minor, if any, impact on our results: First, the majority of the MEE direct short-term response impact on ozone are restricted within 55– magnetic latitudes (Häkkilä et al., **2020**). In our analysis, we utilize ozone data sampled over the entire polar cap regions above geographic latitudes. Second, the major MEE impact are seen at altitudes between 60 and 80 km altitude (Häkkilä et al., **2020**), while in our analysis, we cover altitudes 35–90 km altitudes. Furthermore, considering that our main results (Figures **2** and **3**) shows in general either an agreement throughout the altitude steps, or a disagreement that extends below 60 km altitude. This suggests that in our analysis the MEE contribution, if any, is seen as noise also at the low pfu variability range.

6 Summary and Conclusion

Here, we have validated the capability of WACCM-D to simulate the ion chemistry against the observations from multi-satellites. We derived the atmospheric sensitivity to SPE forcing as a statistical relation between the solar proton flux and the consequent ozone change from satellites. We performed an identical sensitivity analysis for the WACCM-D model and made a one-to-one comparison with the results derived from the satellite observations.

The comparison results show, in general, a good agreement in the nighttime conditions. For daytime, we found a good agreement in the data-sets that include the largest SPEs (max proton flux pfu). For the data-sets including only moderate and minor SPEs (max proton flux pfu), we found WACCM-D to be underestimating the sensitivity compared to the results from satellites.

We conclude that the ion chemistry included in WACCM-D provides a realistic representation of the ozone sensitivity on the EPP driven ionisation, and will therefore provide a good platform for long-term EPP impact studies, e.g., in the context of the CMIP6 simulations.

Acknowledgments